

Multiple-Language Translation System Focusing on Long-distance Medical and Outpatient Services

Rena Aierken and Li Xiao

The Xinjiang Technical Institute of Physics &
Chemistry Academy of Sciences, Urumqi, China
University of Chinese Academy of Science,

Su Sha and Dawa Yidemucao

Key Laboratory of Xinjiang Multi-Language Technology,
Xinjiang University, Urumqi, China
Department of information science and engineering,
Xinjiang University, Urumqi, China

Abstract—For people living in the countryside, an effective long-distance medical and health service is very important. People living in western China, especially, require convenient communication in their native language with doctors working in a modern city. To address this problem, a multiple-language translation system for long-distance medical and outpatient services is discussed. This system initially provides a table containing basic information including disease names and symptoms for different medical classifications, and then translates the sentences selected from the table automatically using a machine translation system. Finally, a PDF file is created for the doctor and the patient. In this paper, the system construction and evaluation of the machine translation are introduced.

Keywords—questionnaire for outpatient cases; Chinese Uyghur language; medical Chinese-Minority language parallel corpus; statistical machine translation.

I. INTRODUCTION

Long-distance medical and health services via internet is extremely beneficial for people living in the countryside. A large percentage of people living in the countryside of Xinjiang cannot speak Mandarin or other languages, while most medical practitioners or doctors do not know the local language. Thus, there is a significant communication-gap between the doctors and the patients. Although a manual translation system could be a possible solution, it costs time and money, besides increasing the burden on patients [1].

A study is going for overcoming the communication problems existing in the medical and health field by building a translation system. The system was built in three steps: first, we built a high-quality Chinese and Uyghur language corpora focusing on the various medical terms in collaboration with medical and health institutions. The corpora were carefully collected by hospital clinics and professionals from medical universities. Next, a set of parallel sentences (PS) including a larger volume of words used in the medical and health field was created. Finally, a translation system that searches the PS and a statistical machine translation (SMT) method were implemented.

When using the translation system, a patient first answers questions in his/her native language. The results of the translation obtained by searching through the PS followed by an SMT are then displayed on the table. Next, the patient selects

the most suitable result from the translation result by clicking on the table buttons. Finally, a PDF file that will be used as an electronic medical record (EMR) is generated by the system automatically.

The remainder of this paper is organized as follows. The related works are presented in section 2, and discuss the overall translation system in section 3. Section 4 shows the experimental setup and analyzes the experimental results. The conclusions and future work will be presented in section 5.

II. RELATED WORKS

Until now, there have been only limited reports on domestic multilingual translation systems for health care research and development. In our previous study, we have reported an electronic medical record translation system that used the strategy of statistical and example-based machine translation for the Chinese and Uyghur languages [2]. The research group of the Ministry of Japan and Ikeda, reported a multi-language support system using a practical text-set for outpatient services [3,4]. They also confirmed that the system performance deteriorates with a larger number of unknown words. Collection of all the relevant words and dialogs in a native language is considerably difficult. In this study, we report a high performance machine translation system focused on the outpatient service. Our system is built based on bilingual parallel sentences in Chinese and Uyghur, relevant to the medical field and the statistical machine translation method.

III. PREPARE YOUR PAPER BEFORE STYLING

A. Questionnaire for the Outpatient Service

An outpatient doctor may provide a patient's condition survey table, as shown in Figure 1, to the patient. The patient answers each question in his/her native language. For demonstration illustration purposes, a Uyghur Latin alphabet-based questionnaire is presented.

To use the system, a patient first answers the question as prompted, entry, depicted in Figure 1 (2) and Figure 1 (3). Next, when the patient clicks on the appropriate buttons, depicted in Figure 1 (2) or Figure 1 (4), the system will display an interface, as shown in Figure 2, automatically, according to patient's input, in two languages (the patient's language and the doctor's language). Column (3) of Figure 2 shows the PS search results and column (4) of Figure 2, shows

the MT result for an entry in column (2) of Figure 2. The PS search results are shown in column (3) of Figure 2, obtained by an N-gram (2 gram or 4 gram) statistical language model and a maximum likelihood criterion. In this case, the system can provide three matching results. If the instance or the input string does not exist in the PS, the system does not display the results. Column (4) of Figure 2 shows the result of the MT. the most suitable result from the translation result by clicking

Common machine translation systems may generate mistranslation results. Therefore, column (5) of Figure 2 is highlighted with a red string display, reminding the user that the machine translation results may have a translation error.

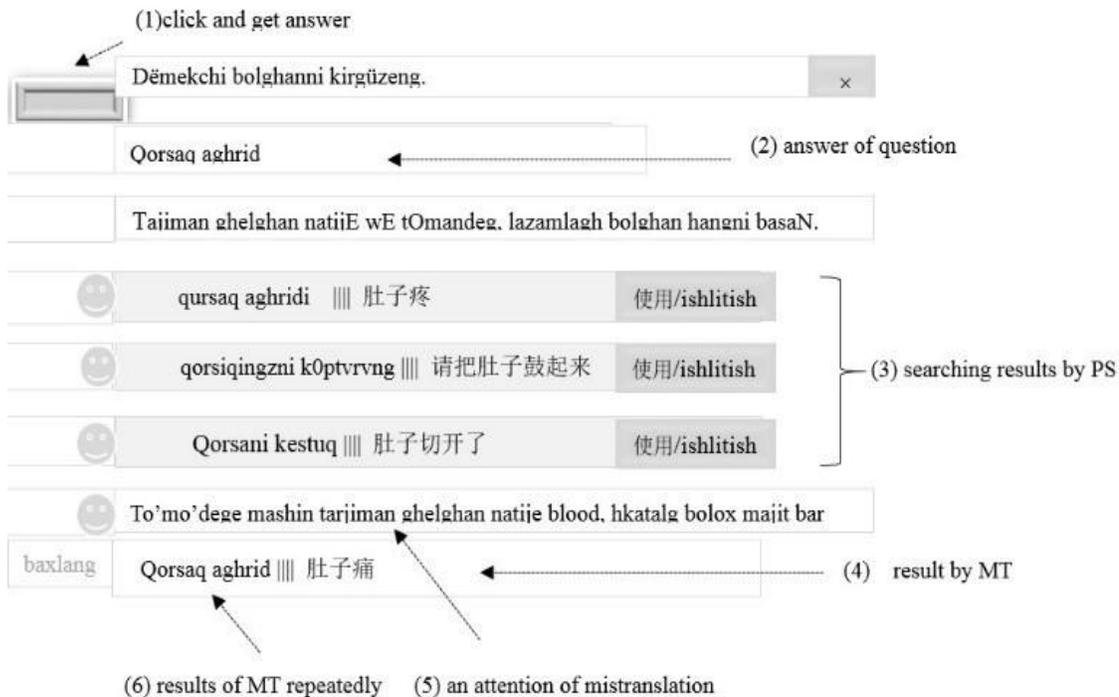


Fig. 1. The complete interface

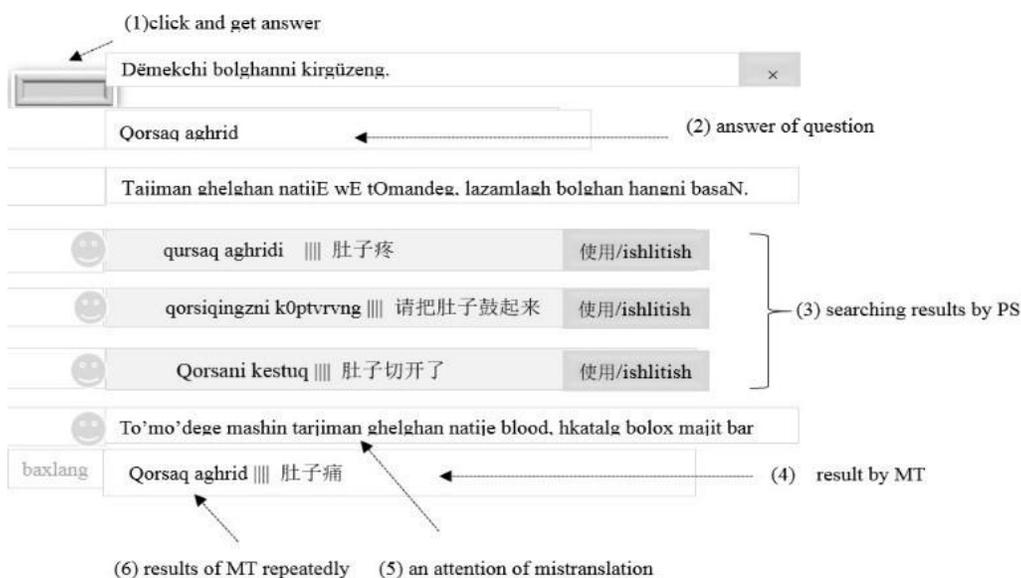


Fig. 2. Translation system

In addition, column (4) of Figure 2 shows the results of the original input strings that are repeatedly translated. Thus, the patient can repeatedly modify the original input, namely column (2) of Figure 2, and choose the most satisfactory translation result to improve the accuracy of the machine translation. The MT provides only one result and the PS gives three, hence, the translation system in Figure 2 can display four different translation results. Finally, the patient can select and click column (3) or column (4) of Figure 2 according to the actual situation, using the right button "use." Then the system prints a PDF file in two languages. Finally, both the doctor and the patient can save the PDF file for electronic medical records.

B. Parallel Sentence and Statistical Machine Translation

This section describes the PS system and the decoder of the machine translation. 240K Chinese sentences including medical, medicine, and drug-names were developed in our previous research. Then, sentences in the Uyghur and Kazak languages corresponding to these Chinese sentences were added manually. Finally, a multi-lingual, aligned medical corpus was created for the system test [5]. We applied a machine translation software based on the general Moses software, developed by the Key Laboratory of multiple language information technology of the Xinjiang University [6,7,8].

IV. EXPERIMENTAL RESULTS

In order to confirm the effectiveness of the questionnaire, we investigated the test results of the PS and the SMT. Then, we tested the performance using an integrated system (PS and SMT).

A. Test Conditions

Ten students, interested in science and engineering, and proficient in Chinese and Uyghur, participated in the experiment. They were asked to enter their own experience of the pathology and relevant medical information. Six students, who specialized in medicine, checked the results of the translation.

To increase the usability of the system, the source language is set to Uyghur and target language is set to Chinese. Additionally,

1) We guaranteed confidentiality in this study. Private information was not revealed.

2) Each participant filled both the investigation form and the system-provided form. These two forms were filled with the same content and in the same order.

3) Taking into account the user's familiarity with the system operation, the two forms were repeated twice. Table I shows examples of the questions. Answers were filled in both a paper-form and a system-provided form. The experiment was carried out with native speakers of the local language (Uyghur). The system assumes that 258 pairs of sentences can be used in the experiment.

TABLE I. QUESTIONNAIRE CONTENTS

No.	So'al sorash mazmi (Question content)	o'al sorash shekli (Question form)
1	qandaq Alamet k�r�lwatidu?	ixtiyari kirg�z�sh
2	qachandin boshlap?	
3	iqki takxurux ainiki bilen tegxarganma?	qemni yaki kunupkini ch�kip kirg�z�ng
4	Qandaq dora yedingiz?	ixtiyari kirg�z�sh
5	dorini hazirmu yemsiz?	

B. Test by Searching the PS only

In this experiment, we investigate two kinds of test results. Figure 3 shows a PS set search result and the MT for question No. 1 in Table I. It is clear that the three results shown in (3) of Figures 2, 3, and 4, by searching the PS alone, are not better than the result of the MT shown in (4) of Figures 2, 3, and 4, for the answer string /axqazan mijaz yaqxiemes (meaning: my stomach is upset) /. This is because the answer string contains commonly used words and does not contain medical terms. We can observe another translation result for question No. 3 in Table I, shown in figure 4. Here, an answer for question No. 3, /iqki takxurux ainiki tegxarixni qlghan/, contains medical terms, hence, the PS search results, (3) of Figures 4 is considerably better than that of the MT, (4) of Figures 3 and 4.

C. Tests using MT only

In this experiment, we assume that the machine translation is an independent implementation. Each sentence that appears in MT is compared with the results. Table II shows the results of the 10 sentences using the PS search. The column, results of the PS (Chinese), includes results of translation by the PS for the source language (Uyghur).

The column, results of the MT, includes results by the MT for the source language (Uyghur). The column, indicating correctness, shows a manual evaluation of the MT for the source language (Uyghur). In this experiment, a total of 20 sentences were selected for the test and 10 sentences were evaluated after excluding repetitive sentences. From Table II, it can be seen that there are four mistranslations among the 10 sentences or four of the translation results do not match the source language. In addition, for a sentence with ID:1 in Table II, although the result using MT and the result of the PS are close in meaning, the result by MT does not answer the question "What are the symptoms?", hence, it is judged as an incorrect result. From the above discussions, it is obvious that

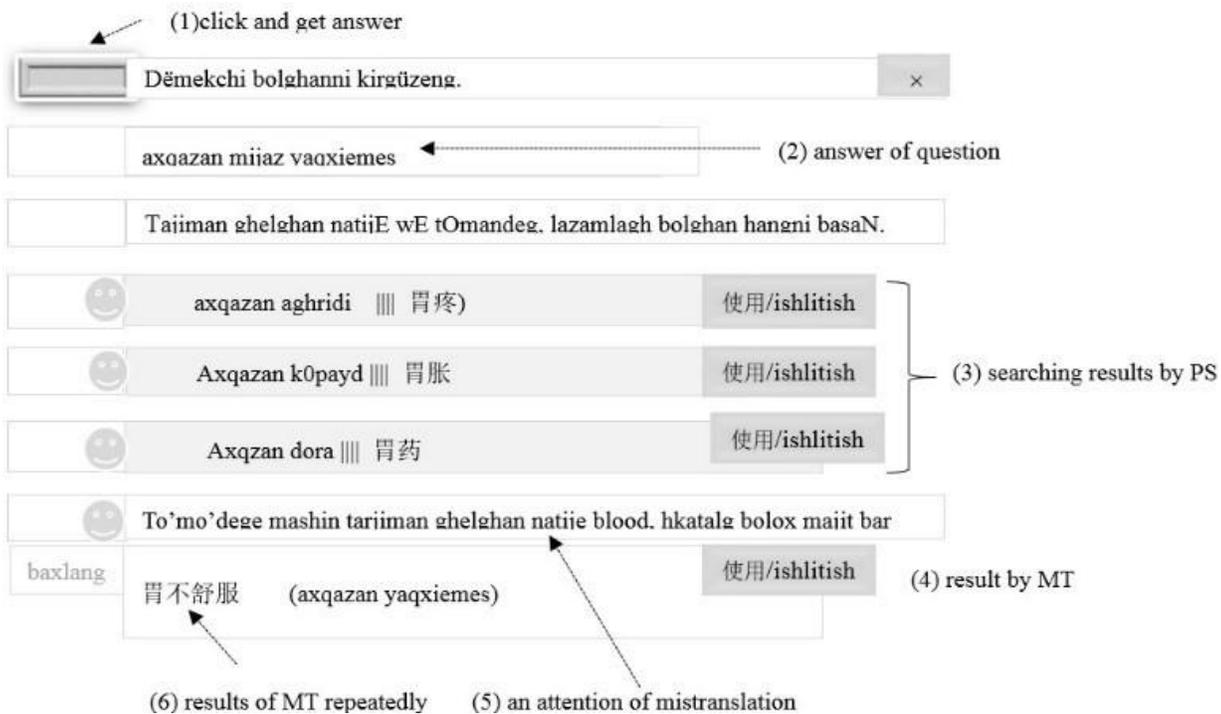


Fig. 3. A translation result for question No. 1

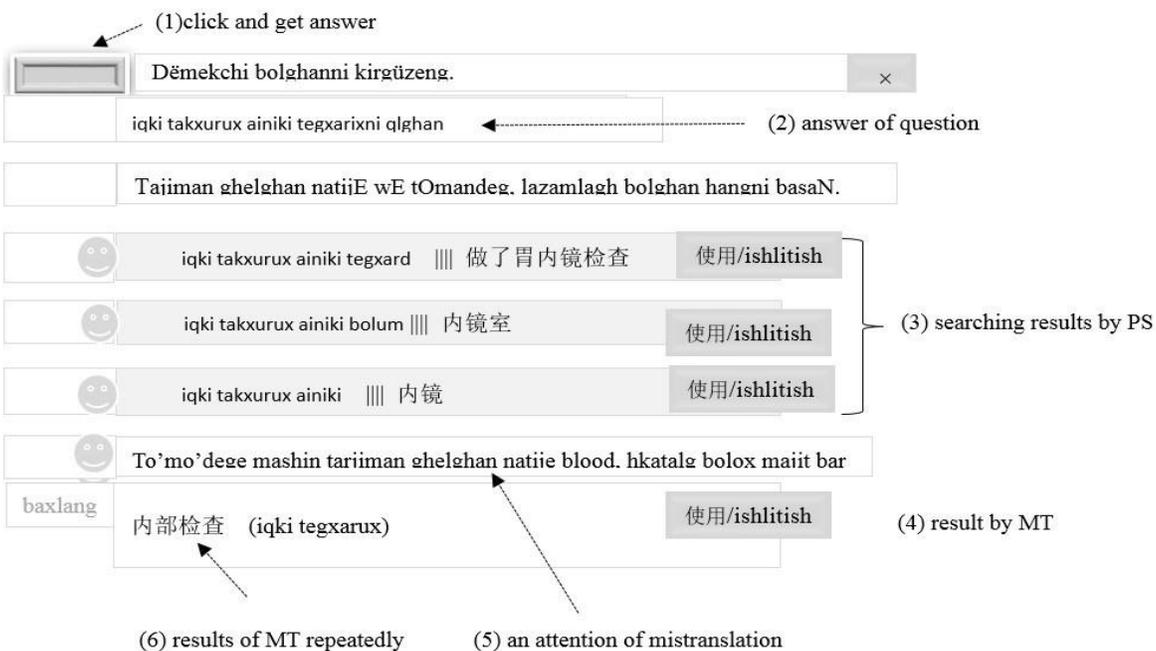


Fig. 4. Translation result for question No. 3

it is difficult to collect all the native language information used in the medical field and the machine translation technique cannot ensure the translation accuracy for the special terms used in the medical and health fields. If the mistranslation by machine translation is to be corrected using the PS search,

then, the proposed method of merging both the PS and the SMT is feasible and complete. A real test EMR is shown in Appendix A.

TABLE II. EXAMPLES OF THE SENTENCES USED IN THE EXPERIMENT

ID	source language (Uyghur)	results of PS (Chinese)	results of MT (Chinese)	correctness
1	Qizitma örlesh	发烧 (fever)	有发热 (have hot)	x
2	Bash aghrish	头疼 (headache)	头疼 (headache)	o
4	Müre sirqirap aghrish	肩膀酸疼 (shoulder ache)	肩膀很疼(shoulder is very painful)	x
8	Zukam	感冒 (have a cold)	感冒 (have a cold)	o
21	Qorsaq aghrish	肚子疼(collywobbles)	肚子疼 (collywobbles)	o
47	Dora rë'aksiye qilish	食物过敏(food allergy)	食品有过敏(food allergies)	x
50	Bash qëyish	头昏(dizzy)	头昏(dizzy)	o
60	Ishtiha yoq	没有食(loss of appetite)	没有食欲 (loss of appetite)	o
72	Nepeste qëlinish	呼吸困难(dyspnea)	呼吸困难 (dyspnea)	o
83	Ashqazan qattiq aghris	胃疼的厉害(have a stomachache)	有时胃疼(sometimes a stomachache)	

V. CONCLUSIONS

In this paper, we have presented a multiple-language translation system that focuses on long-distance medical and outpatient services. The system merges the PS and the SMT approaches and translates the dialogues of the doctors and patients automatically, to create an electronic medical record (EMR). Through experiments, we confirmed that the proposed method is feasible and practical.

In the future, we will further improve the system accuracy and extend the system to service more languages, such as languages along Silk Road.

REFERENCES

- [1] People government of Xinjiang Uyghur Autonomous Region, China. Study and publicize the article implement the second central work conference.<http://www.xj.xinhuanet.com/>
- [2] Dawa Idomucao. A Study on Chinese Minority Medical Document Translation Based on Hybrid-strategy [J]. Journal of Xinjiang University (Natural Science Edition). 2015. Vol. 32(1). 123-128.
- [3] G. Zhenyi. Establishment multi-language Medical Outpatient registered system for foreign patients [J], Journal of the Japan Institute of electronic information and communication. 2009, Vol.192-D No.6 708-718.
- [4] Ikeda, T, Ando, S, Satoh, K Okumura, A. and Watanabe., T, Aoutomatic interpretation system integrating Free-style Sentence translation and Parallel text based Translation[C], Prec.Workshop on speech-to-speech translation 2002,85-92.
- [5] Dawa Idomucao. Research based on multi-language machine translation system with electronic medical record [M], 2012, National Natural Science and Technology Fund: Number:61163030.
- [6] Hieu Hoang and Philipp Koehn. Design of the mooses decoder for statistical machine translation. n Proceedings of ACL Workshop on Software engineering, testing, and quality assurance for NLP, pp. 58–65, 2008.
- [7] Philipp, Koehn, Statistical Machine Translation[M], Uk, Cambridge University,20
- [8] YANG Pan, LI Miao,ZHANG Jian. Chinese-Uyghur ranslation system for phrase-based statical translation [J], Journal of Computer Application 2009 , 29(7),2022-2025.

Appendix A:

医院: دوختۇرخانا
<<电子病历 EMR>>

病类科: بۆلۈم

2015年10月07日

患者姓名: بىمار ئىسمى
出生年月日: تۇغۇلغان يىل ئاي كۈن
民族: مېللىتى
住址: ئادرېس
联系电话: تېلپون

病情调查表: كېسەللىك ئەھۋالى تەكشۈرۈش جەدۋىلى

医生提问?	患者选择回答	机器翻译结果 / ماناسى تەرجىمە / تەتجىسى	医生确认结果 / دوختۇر تەكشۈرۈش نەتىجىسى
1 你有什么症状? / سىزدە قانداق ئەھۋال كۆرۈنىدۇ? دوختۇر سورىغان سوئال?	بىمار تاللىغان جاۋاب يېشىم ئاغرىيدۇ	باش ئاغرىش 头疼	✓
2 什么时候开始? / قاچان باشلانغان? سەزىدە بۇنداق تەبىئىيەت قان بېسىم بارمۇ?	2-3 كۈن	2-3 天 2-3 كۈن	✓
3 你有高血压吗? / سىزدە يۇقىرى قان بېسىم بارمۇ?	بار	有 / بار	✓
4 你现在还服降压药物吗? / سىز ھازىر قان بېسىم چۈشۈرۈش دورىسى ئىچىمىسىز? دورسى ئىچىمىسىز?	يېقىندىن بېرى دورا ئىچىمىگەن	近期没有服药 دورا ئىچىمىگەن	×
5 其他还有什么症状? / يەنە قانداق ئالاھىدە ئەھۋال بار?	دەئىمەت كېسىلمەي بار	دەئىمەت كېسىلگە گىرېتار بولغان 糖尿病	✓
6 对于糖尿病服用了什么药物? / قانداق ئىشلىتىش كېسىلگە نىمە دورا ئىشلىتىۋاتىسىز?	ئىنسىۇلىن	胰岛素 ئىنسىۇلىن	✓
7 你睡眠好吗? / ئۇنىڭغا ياخشى ئەمەس ياكى ياخشى? ياخشىمۇ?	ئانچە ياخشى ئەمەس	不怎么好 ياخشى ئەمەس	✓
8 现在有降压药物吗? / قان بېسىم چۈشۈرۈش دورىسى بارمۇ?	ھازىر يوق	目前没有 ھازىرچە يوق	✓
9 服用几天的降压药物好? / بىرنەچچە كۈنلۈك قان بېسىم چۈشۈرۈش دورىسى ئىشلىتىشىڭىزنىڭ يولۇمۇ? چۈشۈرۈش دورىسى ئىشلىتىشىڭىزنىڭ يولۇمۇ?	يولۇمۇ	吃吧 دورا بەي	✓
10 吃完降压药物再来检查可以吗? / قان بېسىم چۈشۈرۈش دورىسى ئىچىپ بولۇپ يەنە تەكشۈرۈپ باتىشىڭىز قانداق?	يولۇمۇ	可以 يولۇمۇ	✓

医生门诊结果填写: دوختۇر ئامبۇلاتورىيە نەتىجىسى تولدۇرۇلۇشى

1) 患者症状: كېسەل ئەھۋال
2) 门诊结果: دىئاگنوز نەتىجىسى
3) 用药情况: دورا ئىشلىتىش ئەھۋالى
<p>医生签字 (盖章) دوختۇر ئىمزا (تامغا)</p>